practical Machine learning Final Project

Benjie DAI

9/16/2021

Background

This is the final project for the "Pratical Machine Learning" cours in Coursera, In this project, the goal will be to use data from accelerometers on the belt, forearm, arm, and dumbell. and we will try to predict the manner which people did the exercise(into 5 differents classes). source of data: http://groupware.les.inf.puc-rio.br/har.

Read the training data

```
data <- read.csv("pml-training.csv")
summary(data)
```

```
##        X          user_name         raw_timestamp_part_1 raw_timestamp_part_2
##  Min.   :    1   Length:19622       Min.   :1.322e+09    Min.   :   294
##  1st Qu.: 4906   Class :character   1st Qu.:1.323e+09    1st Qu.:252912
##  Median : 9812   Mode  :character   Median :1.323e+09    Median :496380
##  Mean   : 9812                      Mean   :1.323e+09    Mean   :500656
##  3rd Qu.:14717                      3rd Qu.:1.323e+09    3rd Qu.:751891
##  Max.   :19622                      Max.   :1.323e+09    Max.   :998801
##
##  cvtd_timestamp      new_window          num_window      roll_belt
##  Length:19622       Length:19622       Min.   :  1.0   Min.   :-28.90
##  Class :character   Class :character   1st Qu.:222.0   1st Qu.:  1.10
##  Mode  :character   Mode  :character   Median :424.0   Median :113.00
##                                        Mean   :430.6   Mean   : 64.41
##                                        3rd Qu.:644.0   3rd Qu.:123.00
##                                        Max.   :864.0   Max.   :162.00
##
##    pitch_belt          yaw_belt       total_accel_belt kurtosis_roll_belt
##  Min.   :-55.8000   Min.   :-180.00   Min.   : 0.00    Length:19622
##  1st Qu.:  1.7600   1st Qu.: -88.30   1st Qu.: 3.00    Class :character
##  Median :  5.2800   Median : -13.00   Median :17.00    Mode  :character
##  Mean   :  0.3053   Mean   : -11.21   Mean   :11.31
##  3rd Qu.: 14.9000   3rd Qu.:  12.90   3rd Qu.:18.00
##  Max.   : 60.3000   Max.   : 179.00   Max.   :29.00
##
##  kurtosis_picth_belt kurtosis_yaw_belt  skewness_roll_belt skewness_roll_belt.1
##  Length:19622        Length:19622       Length:19622       Length:19622
##  Class :character    Class :character   Class :character   Class :character
```

```
##  Mode  :character    Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  skewness_yaw_belt  max_roll_belt     max_picth_belt  max_yaw_belt
##  Length:19622       Min.   :-94.300   Min.   : 3.00   Length:19622
##  Class :character   1st Qu.:-88.000   1st Qu.: 5.00   Class :character
##  Mode  :character   Median : -5.100   Median :18.00   Mode  :character
##                     Mean   : -6.667   Mean   :12.92
##                     3rd Qu.: 18.500   3rd Qu.:19.00
##                     Max.   :180.000   Max.   :30.00
##                     NA's   :19216     NA's   :19216
##  min_roll_belt     min_pitch_belt  min_yaw_belt       amplitude_roll_belt
##  Min.   :-180.00   Min.   : 0.00   Length:19622       Min.   :  0.000
##  1st Qu.: -88.40   1st Qu.: 3.00   Class :character   1st Qu.:  0.300
##  Median :  -7.85   Median :16.00   Mode  :character   Median :  1.000
##  Mean   : -10.44   Mean   :10.76                      Mean   :  3.769
##  3rd Qu.:   9.05   3rd Qu.:17.00                      3rd Qu.:  2.082
##  Max.   : 173.00   Max.   :23.00                      Max.   :360.000
##  NA's   :19216     NA's   :19216                      NA's   :19216
##  amplitude_pitch_belt amplitude_yaw_belt var_total_accel_belt avg_roll_belt
##  Min.   : 0.000       Length:19622       Min.   : 0.000       Min.   :-27.40
##  1st Qu.: 1.000       Class :character   1st Qu.: 0.100       1st Qu.:  1.10
##  Median : 1.000       Mode  :character   Median : 0.200       Median :116.35
##  Mean   : 2.167                          Mean   : 0.926       Mean   : 68.06
##  3rd Qu.: 2.000                          3rd Qu.: 0.300       3rd Qu.:123.38
##  Max.   :12.000                          Max.   :16.500       Max.   :157.40
##  NA's   :19216                           NA's   :19216        NA's   :19216
##  stddev_roll_belt var_roll_belt     avg_pitch_belt    stddev_pitch_belt
##  Min.   : 0.000   Min.   :  0.000   Min.   :-51.400   Min.   :0.000
##  1st Qu.: 0.200   1st Qu.:  0.000   1st Qu.:  2.025   1st Qu.:0.200
##  Median : 0.400   Median :  0.100   Median :  5.200   Median :0.400
##  Mean   : 1.337   Mean   :  7.699   Mean   :  0.520   Mean   :0.603
##  3rd Qu.: 0.700   3rd Qu.:  0.500   3rd Qu.: 15.775   3rd Qu.:0.700
##  Max.   :14.200   Max.   :200.700   Max.   : 59.700   Max.   :4.000
##  NA's   :19216    NA's   :19216     NA's   :19216     NA's   :19216
##  var_pitch_belt    avg_yaw_belt      stddev_yaw_belt    var_yaw_belt
##  Min.   : 0.000   Min.   :-138.300   Min.   :  0.000   Min.   :    0.000
##  1st Qu.: 0.000   1st Qu.: -88.175   1st Qu.:  0.100   1st Qu.:    0.010
##  Median : 0.100   Median :  -6.550   Median :  0.300   Median :    0.090
##  Mean   : 0.766   Mean   :  -8.831   Mean   :  1.341   Mean   :  107.487
##  3rd Qu.: 0.500   3rd Qu.:  14.125   3rd Qu.:  0.700   3rd Qu.:    0.475
##  Max.   :16.200   Max.   : 173.500   Max.   :176.600   Max.   :31183.240
##  NA's   :19216    NA's   :19216      NA's   :19216     NA's   :19216
##   gyros_belt_x        gyros_belt_y       gyros_belt_z      accel_belt_x
##  Min.   :-1.040000   Min.   :-0.64000   Min.   :-1.4600   Min.   :-120.000
##  1st Qu.:-0.030000   1st Qu.: 0.00000   1st Qu.:-0.2000   1st Qu.: -21.000
##  Median : 0.030000   Median : 0.02000   Median :-0.1000   Median : -15.000
##  Mean   :-0.005592   Mean   : 0.03959   Mean   :-0.1305   Mean   :  -5.595
##  3rd Qu.: 0.110000   3rd Qu.: 0.11000   3rd Qu.:-0.0200   3rd Qu.:  -5.000
##  Max.   : 2.220000   Max.   : 0.64000   Max.   : 1.6200   Max.   :  85.000
##
##   accel_belt_y     accel_belt_z     magnet_belt_x   magnet_belt_y
```

```
##  Min.   :-69.00   Min.   :-275.00   Min.   :-52.0   Min.   :354.0
##  1st Qu.:  3.00   1st Qu.:-162.00   1st Qu.:  9.0   1st Qu.:581.0
##  Median : 35.00   Median :-152.00   Median : 35.0   Median :601.0
##  Mean   : 30.15   Mean   : -72.59   Mean   : 55.6   Mean   :593.7
##  3rd Qu.: 61.00   3rd Qu.:  27.00   3rd Qu.: 59.0   3rd Qu.:610.0
##  Max.   :164.00   Max.   : 105.00   Max.   :485.0   Max.   :673.0
##
##  magnet_belt_z       roll_arm         pitch_arm          yaw_arm
##  Min.   :-623.0   Min.   :-180.00   Min.   :-88.800   Min.   :-180.0000
##  1st Qu.:-375.0   1st Qu.: -31.77   1st Qu.:-25.900   1st Qu.: -43.1000
##  Median :-320.0   Median :   0.00   Median :  0.000   Median :   0.0000
##  Mean   :-345.5   Mean   :  17.83   Mean   : -4.612   Mean   :  -0.6188
##  3rd Qu.:-306.0   3rd Qu.:  77.30   3rd Qu.: 11.200   3rd Qu.:  45.8750
##  Max.   : 293.0   Max.   : 180.00   Max.   : 88.500   Max.   : 180.0000
##
##  total_accel_arm var_accel_arm     avg_roll_arm     stddev_roll_arm
##  Min.   : 1.00   Min.   :  0.00   Min.   :-166.67   Min.   :  0.000
##  1st Qu.:17.00   1st Qu.:  9.03   1st Qu.: -38.37   1st Qu.:  1.376
##  Median :27.00   Median : 40.61   Median :   0.00   Median :  5.702
##  Mean   :25.51   Mean   : 53.23   Mean   :  12.68   Mean   : 11.201
##  3rd Qu.:33.00   3rd Qu.: 75.62   3rd Qu.:  76.33   3rd Qu.: 14.921
##  Max.   :66.00   Max.   :331.70   Max.   : 163.33   Max.   :161.964
##                  NA's   :19216    NA's   :19216     NA's   :19216
##   var_roll_arm       avg_pitch_arm     stddev_pitch_arm var_pitch_arm
##  Min.   :    0.000   Min.   :-81.773   Min.   : 0.000   Min.   :   0.000
##  1st Qu.:    1.898   1st Qu.:-22.770   1st Qu.: 1.642   1st Qu.:   2.697
##  Median :   32.517   Median :  0.000   Median : 8.133   Median :  66.146
##  Mean   :  417.264   Mean   : -4.901   Mean   :10.383   Mean   : 195.864
##  3rd Qu.:  222.647   3rd Qu.:  8.277   3rd Qu.:16.327   3rd Qu.: 266.576
##  Max.   :26232.208   Max.   : 75.659   Max.   :43.412   Max.   :1884.565
##  NA's   :19216       NA's   :19216     NA's   :19216    NA's   :19216
##   avg_yaw_arm       stddev_yaw_arm     var_yaw_arm         gyros_arm_x
##  Min.   :-173.440   Min.   :  0.000   Min.   :    0.000   Min.   :-6.37000
##  1st Qu.: -29.198   1st Qu.:  2.577   1st Qu.:    6.642   1st Qu.:-1.33000
##  Median :   0.000   Median : 16.682   Median :  278.309   Median : 0.08000
##  Mean   :   2.359   Mean   : 22.270   Mean   : 1055.933   Mean   : 0.04277
##  3rd Qu.:  38.185   3rd Qu.: 35.984   3rd Qu.: 1294.850   3rd Qu.: 1.57000
##  Max.   : 152.000   Max.   :177.044   Max.   :31344.568   Max.   : 4.87000
##  NA's   :19216      NA's   :19216     NA's   :19216
##   gyros_arm_y       gyros_arm_z       accel_arm_x       accel_arm_y
##  Min.   :-3.4400   Min.   :-2.3300   Min.   :-404.00   Min.   :-318.0
##  1st Qu.:-0.8000   1st Qu.:-0.0700   1st Qu.:-242.00   1st Qu.: -54.0
##  Median :-0.2400   Median : 0.2300   Median : -44.00   Median :  14.0
##  Mean   :-0.2571   Mean   : 0.2695   Mean   : -60.24   Mean   :  32.6
##  3rd Qu.: 0.1400   3rd Qu.: 0.7200   3rd Qu.:  84.00   3rd Qu.: 139.0
##  Max.   : 2.8400   Max.   : 3.0200   Max.   : 437.00   Max.   : 308.0
##
##   accel_arm_z       magnet_arm_x     magnet_arm_y     magnet_arm_z
##  Min.   :-636.00   Min.   :-584.0   Min.   :-392.0   Min.   :-597.0
##  1st Qu.:-143.00   1st Qu.:-300.0   1st Qu.:  -9.0   1st Qu.: 131.2
##  Median : -47.00   Median : 289.0   Median : 202.0   Median : 444.0
##  Mean   : -71.25   Mean   : 191.7   Mean   : 156.6   Mean   : 306.5
##  3rd Qu.:  23.00   3rd Qu.: 637.0   3rd Qu.: 323.0   3rd Qu.: 545.0
##  Max.   : 292.00   Max.   : 782.0   Max.   : 583.0   Max.   : 694.0
```

```
##
##  kurtosis_roll_arm  kurtosis_picth_arm kurtosis_yaw_arm   skewness_roll_arm
##  Length:19622       Length:19622       Length:19622       Length:19622
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  skewness_pitch_arm skewness_yaw_arm    max_roll_arm     max_picth_arm
##  Length:19622       Length:19622       Min.   :-73.100   Min.   :-173.000
##  Class :character   Class :character   1st Qu.: -0.175   1st Qu.:  -1.975
##  Mode  :character   Mode  :character   Median :  4.950   Median :  23.250
##                                        Mean   : 11.236   Mean   :  35.751
##                                        3rd Qu.: 26.775   3rd Qu.:  95.975
##                                        Max.   : 85.500   Max.   : 180.000
##                                        NA's   :19216     NA's   :19216
##   max_yaw_arm     min_roll_arm    min_pitch_arm      min_yaw_arm
##  Min.   : 4.00   Min.   :-89.10   Min.   :-180.00   Min.   : 1.00
##  1st Qu.:29.00   1st Qu.:-41.98   1st Qu.: -72.62   1st Qu.: 8.00
##  Median :34.00   Median :-22.45   Median : -33.85   Median :13.00
##  Mean   :35.46   Mean   :-21.22   Mean   : -33.92   Mean   :14.66
##  3rd Qu.:41.00   3rd Qu.:  0.00   3rd Qu.:   0.00   3rd Qu.:19.00
##  Max.   :65.00   Max.   : 66.40   Max.   : 152.00   Max.   :38.00
##  NA's   :19216   NA's   :19216    NA's   :19216     NA's   :19216
##  amplitude_roll_arm amplitude_pitch_arm amplitude_yaw_arm roll_dumbbell
##  Min.   :  0.000    Min.   :  0.000     Min.   : 0.00     Min.   :-153.71
##  1st Qu.:  5.425    1st Qu.:  9.925     1st Qu.:13.00     1st Qu.: -18.49
##  Median : 28.450    Median : 54.900     Median :22.00     Median :  48.17
##  Mean   : 32.452    Mean   : 69.677     Mean   :20.79     Mean   :  23.84
##  3rd Qu.: 50.960    3rd Qu.:115.175     3rd Qu.:28.75     3rd Qu.:  67.61
##  Max.   :119.500    Max.   :360.000     Max.   :52.00     Max.   : 153.55
##  NA's   :19216      NA's   :19216       NA's   :19216
##  pitch_dumbbell     yaw_dumbbell      kurtosis_roll_dumbbell
##  Min.   :-149.59   Min.   :-150.871   Length:19622
##  1st Qu.: -40.89   1st Qu.: -77.644   Class :character
##  Median : -20.96   Median :  -3.324   Mode  :character
##  Mean   : -10.78   Mean   :   1.674
##  3rd Qu.:  17.50   3rd Qu.:  79.643
##  Max.   : 149.40   Max.   : 154.952
##
##  kurtosis_picth_dumbbell kurtosis_yaw_dumbbell skewness_roll_dumbbell
##  Length:19622            Length:19622          Length:19622
##  Class :character        Class :character      Class :character
##  Mode  :character        Mode  :character      Mode  :character
##
##
##
##
##  skewness_pitch_dumbbell skewness_yaw_dumbbell max_roll_dumbbell
##  Length:19622            Length:19622          Min.   :-70.10
##  Class :character        Class :character      1st Qu.:-27.15
##  Mode  :character        Mode  :character      Median : 14.85
##                                                Mean   : 13.76
```

```
##                                                 3rd Qu.: 50.58
##                                                 Max.   :137.00
##                                                 NA's   :19216
##  max_picth_dumbbell max_yaw_dumbbell   min_roll_dumbbell min_pitch_dumbbell
##  Min.   :-112.90    Length:19622       Min.   :-149.60   Min.   :-147.00
##  1st Qu.: -66.70    Class :character   1st Qu.: -59.67   1st Qu.: -91.80
##  Median :  40.05    Mode  :character   Median : -43.55   Median : -66.15
##  Mean   :  32.75                       Mean   : -41.24   Mean   : -33.18
##  3rd Qu.: 133.22                       3rd Qu.: -25.20   3rd Qu.:  21.20
##  Max.   : 155.00                       Max.   :  73.20   Max.   : 120.90
##  NA's   :19216                         NA's   :19216     NA's   :19216
##  min_yaw_dumbbell   amplitude_roll_dumbbell amplitude_pitch_dumbbell
##  Length:19622       Min.   :  0.00          Min.   :  0.00
##  Class :character   1st Qu.: 14.97          1st Qu.: 17.06
##  Mode  :character   Median : 35.05          Median : 41.73
##                     Mean   : 55.00          Mean   : 65.93
##                     3rd Qu.: 81.04          3rd Qu.: 99.55
##                     Max.   :256.48          Max.   :273.59
##                     NA's   :19216           NA's   :19216
##  amplitude_yaw_dumbbell total_accel_dumbbell var_accel_dumbbell
##  Length:19622           Min.   : 0.00        Min.   :  0.000
##  Class :character       1st Qu.: 4.00        1st Qu.:  0.378
##  Mode  :character       Median :10.00        Median :  1.000
##                         Mean   :13.72        Mean   :  4.388
##                         3rd Qu.:19.00        3rd Qu.:  3.434
##                         Max.   :58.00        Max.   :230.428
##                                              NA's   :19216
##  avg_roll_dumbbell stddev_roll_dumbbell var_roll_dumbbell  avg_pitch_dumbbell
##  Min.   :-128.96   Min.   :  0.000      Min.   :    0.00   Min.   :-70.73
##  1st Qu.: -12.33   1st Qu.:  4.639      1st Qu.:   21.52   1st Qu.:-42.00
##  Median :  48.23   Median : 12.204      Median :  148.95   Median :-19.91
##  Mean   :  23.86   Mean   : 20.761      Mean   : 1020.27   Mean   :-12.33
##  3rd Qu.:  64.37   3rd Qu.: 26.356      3rd Qu.:  694.65   3rd Qu.: 13.21
##  Max.   : 125.99   Max.   :123.778      Max.   :15321.01   Max.   : 94.28
##  NA's   :19216     NA's   :19216        NA's   :19216      NA's   :19216
##  stddev_pitch_dumbbell var_pitch_dumbbell avg_yaw_dumbbell
##  Min.   : 0.000        Min.   :   0.00    Min.   :-117.950
##  1st Qu.: 3.482        1st Qu.:  12.12    1st Qu.: -76.696
##  Median : 8.089        Median :  65.44    Median :  -4.505
##  Mean   :13.147        Mean   : 350.31    Mean   :   0.202
##  3rd Qu.:19.238        3rd Qu.: 370.11    3rd Qu.:  71.234
##  Max.   :82.680        Max.   :6836.02    Max.   : 134.905
##  NA's   :19216         NA's   :19216      NA's   :19216
##  stddev_yaw_dumbbell var_yaw_dumbbell   gyros_dumbbell_x    gyros_dumbbell_y
##  Min.   :  0.000     Min.   :    0.00   Min.   :-204.0000   Min.   :-2.10000
##  1st Qu.:  3.885     1st Qu.:   15.09   1st Qu.:  -0.0300   1st Qu.:-0.14000
##  Median : 10.264     Median :  105.35   Median :   0.1300   Median : 0.03000
##  Mean   : 16.647     Mean   :  589.84   Mean   :   0.1611   Mean   : 0.04606
##  3rd Qu.: 24.674     3rd Qu.:  608.79   3rd Qu.:   0.3500   3rd Qu.: 0.21000
##  Max.   :107.088     Max.   :11467.91   Max.   :   2.2200   Max.   :52.00000
##  NA's   :19216       NA's   :19216
##  gyros_dumbbell_z  accel_dumbbell_x  accel_dumbbell_y  accel_dumbbell_z
##  Min.   : -2.380   Min.   :-419.00   Min.   :-189.00   Min.   :-334.00
##  1st Qu.: -0.310   1st Qu.: -50.00   1st Qu.:  -8.00   1st Qu.:-142.00
```

```
##  Median : -0.130   Median :  -8.00   Median :  41.50   Median :  -1.00
##  Mean   : -0.129   Mean   : -28.62   Mean   :  52.63   Mean   : -38.32
##  3rd Qu.:  0.030   3rd Qu.:  11.00   3rd Qu.: 111.00   3rd Qu.:  38.00
##  Max.   :317.000   Max.   : 235.00   Max.   : 315.00   Max.   : 318.00
##
##  magnet_dumbbell_x magnet_dumbbell_y magnet_dumbbell_z  roll_forearm
##  Min.   :-643.0    Min.   :-3600     Min.   :-262.00   Min.   :-180.0000
##  1st Qu.:-535.0    1st Qu.:  231     1st Qu.: -45.00   1st Qu.:  -0.7375
##  Median :-479.0    Median :  311     Median :  13.00   Median :  21.7000
##  Mean   :-328.5    Mean   :  221     Mean   :  46.05   Mean   :  33.8265
##  3rd Qu.:-304.0    3rd Qu.:  390     3rd Qu.:  95.00   3rd Qu.: 140.0000
##  Max.   : 592.0    Max.   :  633     Max.   : 452.00   Max.   : 180.0000
##
##  pitch_forearm     yaw_forearm      kurtosis_roll_forearm
##  Min.   :-72.50   Min.   :-180.00   Length:19622
##  1st Qu.:  0.00   1st Qu.: -68.60   Class :character
##  Median :  9.24   Median :   0.00   Mode  :character
##  Mean   : 10.71   Mean   :  19.21
##  3rd Qu.: 28.40   3rd Qu.: 110.00
##  Max.   : 89.80   Max.   : 180.00
##
##  kurtosis_picth_forearm kurtosis_yaw_forearm skewness_roll_forearm
##  Length:19622           Length:19622         Length:19622
##  Class :character       Class :character     Class :character
##  Mode  :character       Mode  :character     Mode  :character
##
##
##
##
##  skewness_pitch_forearm skewness_yaw_forearm max_roll_forearm max_picth_forearm
##  Length:19622           Length:19622         Min.   :-66.60   Min.   :-151.00
##  Class :character       Class :character     1st Qu.:  0.00   1st Qu.:   0.00
##  Mode  :character       Mode  :character     Median : 26.80   Median : 113.00
##                                              Mean   : 24.49   Mean   :  81.49
##                                              3rd Qu.: 45.95   3rd Qu.: 174.75
##                                              Max.   : 89.80   Max.   : 180.00
##                                              NA's   :19216    NA's   :19216
##  max_yaw_forearm    min_roll_forearm  min_pitch_forearm min_yaw_forearm
##  Length:19622       Min.   :-72.500   Min.   :-180.00   Length:19622
##  Class :character   1st Qu.: -6.075   1st Qu.:-175.00   Class :character
##  Mode  :character   Median :  0.000   Median : -61.00   Mode  :character
##                     Mean   : -0.167   Mean   : -57.57
##                     3rd Qu.: 12.075   3rd Qu.:   0.00
##                     Max.   : 62.100   Max.   : 167.00
##                     NA's   :19216     NA's   :19216
##  amplitude_roll_forearm amplitude_pitch_forearm amplitude_yaw_forearm
##  Min.   :  0.000        Min.   :  0.0           Length:19622
##  1st Qu.:  1.125        1st Qu.:  2.0           Class :character
##  Median : 17.770        Median : 83.7           Mode  :character
##  Mean   : 24.653        Mean   :139.1
##  3rd Qu.: 39.875        3rd Qu.:350.0
##  Max.   :126.000        Max.   :360.0
##  NA's   :19216          NA's   :19216
##  total_accel_forearm var_accel_forearm avg_roll_forearm   stddev_roll_forearm
```

```
##  Min.   :  0.00      Min.   :  0.000   Min.   :-177.234   Min.   :  0.000
##  1st Qu.: 29.00      1st Qu.:  6.759   1st Qu.:  -0.909   1st Qu.:  0.428
##  Median : 36.00      Median : 21.165   Median :  11.172   Median :  8.030
##  Mean   : 34.72      Mean   : 33.502   Mean   :  33.165   Mean   : 41.986
##  3rd Qu.: 41.00      3rd Qu.: 51.240   3rd Qu.: 107.132   3rd Qu.: 85.373
##  Max.   :108.00      Max.   :172.606   Max.   : 177.256   Max.   :179.171
##                      NA's   :19216     NA's   :19216      NA's   :19216
##  var_roll_forearm   avg_pitch_forearm stddev_pitch_forearm var_pitch_forearm
##  Min.   :    0.00   Min.   :-68.17    Min.   : 0.000       Min.   :   0.000
##  1st Qu.:    0.18   1st Qu.:  0.00    1st Qu.: 0.336       1st Qu.:   0.113
##  Median :   64.48   Median : 12.02    Median : 5.516       Median :  30.425
##  Mean   : 5274.10   Mean   : 11.79    Mean   : 7.977       Mean   : 139.593
##  3rd Qu.: 7289.08   3rd Qu.: 28.48    3rd Qu.:12.866       3rd Qu.: 165.532
##  Max.   :32102.24   Max.   : 72.09    Max.   :47.745       Max.   :2279.617
##  NA's   :19216      NA's   :19216     NA's   :19216        NA's   :19216
##  avg_yaw_forearm   stddev_yaw_forearm var_yaw_forearm    gyros_forearm_x
##  Min.   :-155.06   Min.   :  0.000    Min.   :    0.00   Min.   :-22.000
##  1st Qu.: -26.26   1st Qu.:  0.524    1st Qu.:    0.27   1st Qu.: -0.220
##  Median :   0.00   Median : 24.743    Median :  612.21   Median :  0.050
##  Mean   :  18.00   Mean   : 44.854    Mean   : 4639.85   Mean   :  0.158
##  3rd Qu.:  85.79   3rd Qu.: 85.817    3rd Qu.: 7368.41   3rd Qu.:  0.560
##  Max.   : 169.24   Max.   :197.508    Max.   :39009.33   Max.   :  3.970
##  NA's   :19216     NA's   :19216      NA's   :19216
##  gyros_forearm_y     gyros_forearm_z    accel_forearm_x   accel_forearm_y
##  Min.   : -7.02000   Min.   : -8.0900   Min.   :-498.00   Min.   :-632.0
##  1st Qu.: -1.46000   1st Qu.: -0.1800   1st Qu.:-178.00   1st Qu.:  57.0
##  Median :  0.03000   Median :  0.0800   Median : -57.00   Median : 201.0
##  Mean   :  0.07517   Mean   :  0.1512   Mean   : -61.65   Mean   : 163.7
##  3rd Qu.:  1.62000   3rd Qu.:  0.4900   3rd Qu.:  76.00   3rd Qu.: 312.0
##  Max.   :311.00000   Max.   :231.0000   Max.   : 477.00   Max.   : 923.0
##
##  accel_forearm_z   magnet_forearm_x  magnet_forearm_y magnet_forearm_z
##  Min.   :-446.00   Min.   :-1280.0   Min.   :-896.0   Min.   :-973.0
##  1st Qu.:-182.00   1st Qu.: -616.0   1st Qu.:   2.0   1st Qu.: 191.0
##  Median : -39.00   Median : -378.0   Median : 591.0   Median : 511.0
##  Mean   : -55.29   Mean   : -312.6   Mean   : 380.1   Mean   : 393.6
##  3rd Qu.:  26.00   3rd Qu.:  -73.0   3rd Qu.: 737.0   3rd Qu.: 653.0
##  Max.   : 291.00   Max.   :  672.0   Max.   :1480.0   Max.   :1090.0
##
##     classe
##  Length:19622
##  Class :character
##  Mode  :character
##
##
##
##
```

```
str(data)
```

```
## 'data.frame':    19622 obs. of  160 variables:
##  $ X                       : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ user_name               : chr  "carlitos" "carlitos" "carlitos" "carlitos" ...
##  $ raw_timestamp_part_1    : int  1323084231 1323084231 1323084231 1323084232 1323084232 1323084232 1
```

```
##  $ raw_timestamp_part_2    : int  788290 808298 820366 120339 196328 304277 368296 440390 484323 484
##  $ cvtd_timestamp          : chr  "05/12/2011 11:23" "05/12/2011 11:23" "05/12/2011 11:23" "05/12/20
##  $ new_window              : chr  "no" "no" "no" "no" ...
##  $ num_window              : int  11 11 11 12 12 12 12 12 12 12 ...
##  $ roll_belt               : num  1.41 1.41 1.42 1.48 1.48 1.45 1.42 1.42 1.43 1.45 ...
##  $ pitch_belt              : num  8.07 8.07 8.07 8.05 8.07 8.06 8.09 8.13 8.16 8.17 ...
##  $ yaw_belt                : num  -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 ...
##  $ total_accel_belt        : int  3 3 3 3 3 3 3 3 3 3 ...
##  $ kurtosis_roll_belt      : chr  "" "" "" "" ...
##  $ kurtosis_picth_belt     : chr  "" "" "" "" ...
##  $ kurtosis_yaw_belt       : chr  "" "" "" "" ...
##  $ skewness_roll_belt      : chr  "" "" "" "" ...
##  $ skewness_roll_belt.1    : chr  "" "" "" "" ...
##  $ skewness_yaw_belt       : chr  "" "" "" "" ...
##  $ max_roll_belt           : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ max_picth_belt          : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ max_yaw_belt            : chr  "" "" "" "" ...
##  $ min_roll_belt           : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ min_pitch_belt          : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ min_yaw_belt            : chr  "" "" "" "" ...
##  $ amplitude_roll_belt     : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ amplitude_pitch_belt    : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ amplitude_yaw_belt      : chr  "" "" "" "" ...
##  $ var_total_accel_belt    : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ avg_roll_belt           : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ stddev_roll_belt        : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ var_roll_belt           : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ avg_pitch_belt          : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ stddev_pitch_belt       : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ var_pitch_belt          : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ avg_yaw_belt            : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ stddev_yaw_belt         : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ var_yaw_belt            : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ gyros_belt_x            : num  0 0.02 0 0.02 0.02 0.02 0.02 0.02 0.02 0.03 ...
##  $ gyros_belt_y            : num  0 0 0 0 0.02 0 0 0 0 0 ...
##  $ gyros_belt_z            : num  -0.02 -0.02 -0.02 -0.03 -0.02 -0.02 -0.02 -0.02 -0.02 0 ...
##  $ accel_belt_x            : int  -21 -22 -20 -22 -21 -21 -22 -22 -20 -21 ...
##  $ accel_belt_y            : int  4 4 5 3 2 4 3 4 2 4 ...
##  $ accel_belt_z            : int  22 22 23 21 24 21 21 21 24 22 ...
##  $ magnet_belt_x           : int  -3 -7 -2 -6 -6 0 -4 -2 1 -3 ...
##  $ magnet_belt_y           : int  599 608 600 604 600 603 599 603 602 609 ...
##  $ magnet_belt_z           : int  -313 -311 -305 -310 -302 -312 -311 -313 -312 -308 ...
##  $ roll_arm                : num  -128 -128 -128 -128 -128 -128 -128 -128 -128 -128 ...
##  $ pitch_arm               : num  22.5 22.5 22.5 22.1 22.1 22 21.9 21.8 21.7 21.6 ...
##  $ yaw_arm                 : num  -161 -161 -161 -161 -161 -161 -161 -161 -161 -161 ...
##  $ total_accel_arm         : int  34 34 34 34 34 34 34 34 34 34 ...
##  $ var_accel_arm           : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ avg_roll_arm            : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ stddev_roll_arm         : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ var_roll_arm            : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ avg_pitch_arm           : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ stddev_pitch_arm        : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ var_pitch_arm           : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ avg_yaw_arm             : num  NA NA NA NA NA NA NA NA NA NA ...
```

```
##  $ stddev_yaw_arm          : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ var_yaw_arm             : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ gyros_arm_x             : num  0 0.02 0.02 0.02 0 0.02 0 0.02 0.02 0.02 ...
##  $ gyros_arm_y             : num  0 -0.02 -0.02 -0.03 -0.03 -0.03 -0.03 -0.02 -0.03 -0.03 ...
##  $ gyros_arm_z             : num  -0.02 -0.02 -0.02 0.02 0 0 0 0 -0.02 -0.02 ...
##  $ accel_arm_x             : int  -288 -290 -289 -289 -289 -289 -289 -289 -288 -288 ...
##  $ accel_arm_y             : int  109 110 110 111 111 111 111 111 109 110 ...
##  $ accel_arm_z             : int  -123 -125 -126 -123 -123 -122 -125 -124 -122 -124 ...
##  $ magnet_arm_x            : int  -368 -369 -368 -372 -374 -369 -373 -372 -369 -376 ...
##  $ magnet_arm_y            : int  337 337 344 344 337 342 336 338 341 334 ...
##  $ magnet_arm_z            : int  516 513 513 512 506 513 509 510 518 516 ...
##  $ kurtosis_roll_arm       : chr  "" "" "" "" ...
##  $ kurtosis_picth_arm      : chr  "" "" "" "" ...
##  $ kurtosis_yaw_arm        : chr  "" "" "" "" ...
##  $ skewness_roll_arm       : chr  "" "" "" "" ...
##  $ skewness_pitch_arm      : chr  "" "" "" "" ...
##  $ skewness_yaw_arm        : chr  "" "" "" "" ...
##  $ max_roll_arm            : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ max_picth_arm           : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ max_yaw_arm             : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ min_roll_arm            : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ min_pitch_arm           : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ min_yaw_arm             : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ amplitude_roll_arm      : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ amplitude_pitch_arm     : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ amplitude_yaw_arm       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ roll_dumbbell           : num  13.1 13.1 12.9 13.4 13.4 ...
##  $ pitch_dumbbell          : num  -70.5 -70.6 -70.3 -70.4 -70.4 ...
##  $ yaw_dumbbell            : num  -84.9 -84.7 -85.1 -84.9 -84.9 ...
##  $ kurtosis_roll_dumbbell  : chr  "" "" "" "" ...
##  $ kurtosis_picth_dumbbell : chr  "" "" "" "" ...
##  $ kurtosis_yaw_dumbbell   : chr  "" "" "" "" ...
##  $ skewness_roll_dumbbell  : chr  "" "" "" "" ...
##  $ skewness_pitch_dumbbell : chr  "" "" "" "" ...
##  $ skewness_yaw_dumbbell   : chr  "" "" "" "" ...
##  $ max_roll_dumbbell       : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ max_picth_dumbbell      : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ max_yaw_dumbbell        : chr  "" "" "" "" ...
##  $ min_roll_dumbbell       : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ min_pitch_dumbbell      : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ min_yaw_dumbbell        : chr  "" "" "" "" ...
##  $ amplitude_roll_dumbbell : num  NA NA NA NA NA NA NA NA NA NA ...
##    [list output truncated]
```

```
table(data$classe)
```

```
##
##    A    B    C    D    E
## 5580 3797 3422 3216 3607
```

check the not Na Ratio

```
nacount <- apply(data,2, function(x) {sum(!is.na(x))/length(x)})
nacount
```

```
##                        X                user_name     raw_timestamp_part_1
##               1.00000000               1.00000000               1.00000000
##     raw_timestamp_part_2           cvtd_timestamp               new_window
##               1.00000000               1.00000000               1.00000000
##               num_window              roll_belt               pitch_belt
##               1.00000000               1.00000000               1.00000000
##                yaw_belt         total_accel_belt       kurtosis_roll_belt
##               1.00000000               1.00000000               1.00000000
##      kurtosis_picth_belt        kurtosis_yaw_belt       skewness_roll_belt
##               1.00000000               1.00000000               1.00000000
##     skewness_roll_belt.1        skewness_yaw_belt            max_roll_belt
##               1.00000000               1.00000000               0.02069106
##           max_picth_belt             max_yaw_belt            min_roll_belt
##               0.02069106               1.00000000               0.02069106
##           min_pitch_belt             min_yaw_belt      amplitude_roll_belt
##               0.02069106               1.00000000               0.02069106
##     amplitude_pitch_belt       amplitude_yaw_belt     var_total_accel_belt
##               0.02069106               1.00000000               0.02069106
##            avg_roll_belt          stddev_roll_belt            var_roll_belt
##               0.02069106               0.02069106               0.02069106
##           avg_pitch_belt         stddev_pitch_belt           var_pitch_belt
##               0.02069106               0.02069106               0.02069106
##             avg_yaw_belt           stddev_yaw_belt             var_yaw_belt
##               0.02069106               0.02069106               0.02069106
##             gyros_belt_x             gyros_belt_y             gyros_belt_z
##               1.00000000               1.00000000               1.00000000
##             accel_belt_x             accel_belt_y             accel_belt_z
##               1.00000000               1.00000000               1.00000000
##            magnet_belt_x            magnet_belt_y            magnet_belt_z
##               1.00000000               1.00000000               1.00000000
##                roll_arm                 pitch_arm                   yaw_arm
##               1.00000000               1.00000000               1.00000000
##          total_accel_arm            var_accel_arm            avg_roll_arm
##               1.00000000               0.02069106               0.02069106
##           stddev_roll_arm            var_roll_arm            avg_pitch_arm
##               0.02069106               0.02069106               0.02069106
##          stddev_pitch_arm           var_pitch_arm             avg_yaw_arm
##               0.02069106               0.02069106               0.02069106
##           stddev_yaw_arm             var_yaw_arm             gyros_arm_x
##               0.02069106               0.02069106               1.00000000
##              gyros_arm_y              gyros_arm_z              accel_arm_x
##               1.00000000               1.00000000               1.00000000
##              accel_arm_y              accel_arm_z             magnet_arm_x
##               1.00000000               1.00000000               1.00000000
##             magnet_arm_y             magnet_arm_z        kurtosis_roll_arm
##               1.00000000               1.00000000               1.00000000
##        kurtosis_picth_arm         kurtosis_yaw_arm         skewness_roll_arm
##               1.00000000               1.00000000               1.00000000
```

```
##       skewness_pitch_arm         skewness_yaw_arm              max_roll_arm
##               1.00000000               1.00000000               0.02069106
##            max_picth_arm             max_yaw_arm              min_roll_arm
##               0.02069106               0.02069106               0.02069106
##            min_pitch_arm             min_yaw_arm       amplitude_roll_arm
##               0.02069106               0.02069106               0.02069106
##      amplitude_pitch_arm        amplitude_yaw_arm            roll_dumbbell
##               0.02069106               0.02069106               1.00000000
##           pitch_dumbbell             yaw_dumbbell   kurtosis_roll_dumbbell
##               1.00000000               1.00000000               1.00000000
##  kurtosis_picth_dumbbell    kurtosis_yaw_dumbbell   skewness_roll_dumbbell
##               1.00000000               1.00000000               1.00000000
##  skewness_pitch_dumbbell    skewness_yaw_dumbbell        max_roll_dumbbell
##               1.00000000               1.00000000               0.02069106
##       max_picth_dumbbell         max_yaw_dumbbell        min_roll_dumbbell
##               0.02069106               1.00000000               0.02069106
##       min_pitch_dumbbell         min_yaw_dumbbell  amplitude_roll_dumbbell
##               0.02069106               1.00000000               0.02069106
## amplitude_pitch_dumbbell   amplitude_yaw_dumbbell     total_accel_dumbbell
##               0.02069106               1.00000000               1.00000000
##       var_accel_dumbbell        avg_roll_dumbbell     stddev_roll_dumbbell
##               0.02069106               0.02069106               0.02069106
##        var_roll_dumbbell       avg_pitch_dumbbell    stddev_pitch_dumbbell
##               0.02069106               0.02069106               0.02069106
##       var_pitch_dumbbell         avg_yaw_dumbbell      stddev_yaw_dumbbell
##               0.02069106               0.02069106               0.02069106
##         var_yaw_dumbbell         gyros_dumbbell_x         gyros_dumbbell_y
##               0.02069106               1.00000000               1.00000000
##         gyros_dumbbell_z         accel_dumbbell_x         accel_dumbbell_y
##               1.00000000               1.00000000               1.00000000
##         accel_dumbbell_z        magnet_dumbbell_x        magnet_dumbbell_y
##               1.00000000               1.00000000               1.00000000
##        magnet_dumbbell_z             roll_forearm            pitch_forearm
##               1.00000000               1.00000000               1.00000000
##             yaw_forearm    kurtosis_roll_forearm   kurtosis_picth_forearm
##               1.00000000               1.00000000               1.00000000
##     kurtosis_yaw_forearm    skewness_roll_forearm   skewness_pitch_forearm
##               1.00000000               1.00000000               1.00000000
##     skewness_yaw_forearm         max_roll_forearm        max_picth_forearm
##               1.00000000               0.02069106               0.02069106
##          max_yaw_forearm         min_roll_forearm        min_pitch_forearm
##               1.00000000               0.02069106               0.02069106
##          min_yaw_forearm   amplitude_roll_forearm  amplitude_pitch_forearm
##               1.00000000               0.02069106               0.02069106
##    amplitude_yaw_forearm      total_accel_forearm        var_accel_forearm
##               1.00000000               1.00000000               0.02069106
##         avg_roll_forearm      stddev_roll_forearm         var_roll_forearm
##               0.02069106               0.02069106               0.02069106
##        avg_pitch_forearm     stddev_pitch_forearm        var_pitch_forearm
##               0.02069106               0.02069106               0.02069106
##          avg_yaw_forearm       stddev_yaw_forearm          var_yaw_forearm
##               0.02069106               0.02069106               0.02069106
##          gyros_forearm_x          gyros_forearm_y          gyros_forearm_z
##               1.00000000               1.00000000               1.00000000
```

```
##          accel_forearm_x          accel_forearm_y          accel_forearm_z
##               1.00000000               1.00000000               1.00000000
##         magnet_forearm_x         magnet_forearm_y         magnet_forearm_z
##               1.00000000               1.00000000               1.00000000
##                   classe
##               1.00000000
```

Identify the columns to keep with criteria : not NA value ratio > 80%

```
keepcol=c()
counter=1
for (i in (1:length(nacount))) {
  if(nacount[[i]]>0.8){
    keepcol[counter] <- names(nacount[i])
    counter <- counter+1
  }
}
data1 <- subset(data,select =keepcol)
```

drop the useless information

```
useless <- c("X","user_name","raw_timestamp_part_1","raw_timestamp_part_2","cvtd_timestamp",
             "new_window","num_window")
data2 <- data1[,!(names(data1) %in% useless)]
```

drop the columns where contains 90% of emply data

```
emptydata <- c("kurtosis_roll_belt","kurtosis_picth_belt","kurtosis_picth_belt","kurtosis_yaw_belt",
               "skewness_roll_belt","skewness_roll_belt.1","skewness_yaw_belt",
              "max_yaw_belt","min_yaw_belt","amplitude_yaw_belt","kurtosis_roll_arm",
              "kurtosis_picth_arm","kurtosis_yaw_arm","skewness_roll_arm","skewness_pitch_arm",
              "skewness_yaw_arm","kurtosis_roll_dumbbell","kurtosis_picth_dumbbell",
              "kurtosis_yaw_dumbbell","skewness_roll_dumbbell","skewness_pitch_dumbbell",
              "skewness_yaw_dumbbell","max_yaw_dumbbell","min_yaw_dumbbell","amplitude_yaw_dumbbell",
              "kurtosis_roll_forearm","kurtosis_picth_forearm","kurtosis_yaw_forearm",
              "skewness_roll_forearm","skewness_pitch_forearm","skewness_yaw_forearm",
              "max_yaw_forearm","min_yaw_forearm","amplitude_yaw_forearm")
```

drop all columns need to be removed

```
data3 <- data2[,!names(data2) %in% emptydata]
summary(data3)
```

```
##    roll_belt        pitch_belt          yaw_belt       total_accel_belt
##  Min.   :-28.90   Min.   :-55.8000   Min.   :-180.00   Min.   : 0.00
```

```
##  1st Qu.:  1.10   1st Qu.:  1.7600   1st Qu.: -88.30   1st Qu.: 3.00
##  Median :113.00   Median :  5.2800   Median : -13.00   Median :17.00
##  Mean   : 64.41   Mean   :  0.3053   Mean   : -11.21   Mean   :11.31
##  3rd Qu.:123.00   3rd Qu.: 14.9000   3rd Qu.:  12.90   3rd Qu.:18.00
##  Max.   :162.00   Max.   : 60.3000   Max.   : 179.00   Max.   :29.00
##   gyros_belt_x        gyros_belt_y       gyros_belt_z      accel_belt_x
##  Min.   :-1.040000   Min.   :-0.64000   Min.   :-1.4600   Min.   :-120.000
##  1st Qu.:-0.030000   1st Qu.: 0.00000   1st Qu.:-0.2000   1st Qu.: -21.000
##  Median : 0.030000   Median : 0.02000   Median :-0.1000   Median : -15.000
##  Mean   :-0.005592   Mean   : 0.03959   Mean   :-0.1305   Mean   :  -5.595
##  3rd Qu.: 0.110000   3rd Qu.: 0.11000   3rd Qu.:-0.0200   3rd Qu.:  -5.000
##  Max.   : 2.220000   Max.   : 0.64000   Max.   : 1.6200   Max.   :  85.000
##   accel_belt_y     accel_belt_z     magnet_belt_x   magnet_belt_y
##  Min.   :-69.00   Min.   :-275.00   Min.   :-52.0   Min.   :354.0
##  1st Qu.:  3.00   1st Qu.:-162.00   1st Qu.:  9.0   1st Qu.:581.0
##  Median : 35.00   Median :-152.00   Median : 35.0   Median :601.0
##  Mean   : 30.15   Mean   : -72.59   Mean   : 55.6   Mean   :593.7
##  3rd Qu.: 61.00   3rd Qu.:  27.00   3rd Qu.: 59.0   3rd Qu.:610.0
##  Max.   :164.00   Max.   : 105.00   Max.   :485.0   Max.   :673.0
##  magnet_belt_z       roll_arm         pitch_arm          yaw_arm
##  Min.   :-623.0   Min.   :-180.00   Min.   :-88.800   Min.   :-180.0000
##  1st Qu.:-375.0   1st Qu.: -31.77   1st Qu.:-25.900   1st Qu.: -43.1000
##  Median :-320.0   Median :   0.00   Median :  0.000   Median :   0.0000
##  Mean   :-345.5   Mean   :  17.83   Mean   : -4.612   Mean   :  -0.6188
##  3rd Qu.:-306.0   3rd Qu.:  77.30   3rd Qu.: 11.200   3rd Qu.:  45.8750
##  Max.   : 293.0   Max.   : 180.00   Max.   : 88.500   Max.   : 180.0000
##  total_accel_arm  gyros_arm_x        gyros_arm_y       gyros_arm_z
##  Min.   : 1.00   Min.   :-6.37000   Min.   :-3.4400   Min.   :-2.3300
##  1st Qu.:17.00   1st Qu.:-1.33000   1st Qu.:-0.8000   1st Qu.:-0.0700
##  Median :27.00   Median : 0.08000   Median :-0.2400   Median : 0.2300
##  Mean   :25.51   Mean   : 0.04277   Mean   :-0.2571   Mean   : 0.2695
##  3rd Qu.:33.00   3rd Qu.: 1.57000   3rd Qu.: 0.1400   3rd Qu.: 0.7200
##  Max.   :66.00   Max.   : 4.87000   Max.   : 2.8400   Max.   : 3.0200
##   accel_arm_x       accel_arm_y      accel_arm_z       magnet_arm_x
##  Min.   :-404.00   Min.   :-318.0   Min.   :-636.00   Min.   :-584.0
##  1st Qu.:-242.00   1st Qu.: -54.0   1st Qu.:-143.00   1st Qu.:-300.0
##  Median : -44.00   Median :  14.0   Median : -47.00   Median : 289.0
##  Mean   : -60.24   Mean   :  32.6   Mean   : -71.25   Mean   : 191.7
##  3rd Qu.:  84.00   3rd Qu.: 139.0   3rd Qu.:  23.00   3rd Qu.: 637.0
##  Max.   : 437.00   Max.   : 308.0   Max.   : 292.00   Max.   : 782.0
##   magnet_arm_y     magnet_arm_z    roll_dumbbell     pitch_dumbbell
##  Min.   :-392.0   Min.   :-597.0   Min.   :-153.71   Min.   :-149.59
##  1st Qu.:  -9.0   1st Qu.: 131.2   1st Qu.: -18.49   1st Qu.: -40.89
##  Median : 202.0   Median : 444.0   Median :  48.17   Median : -20.96
##  Mean   : 156.6   Mean   : 306.5   Mean   :  23.84   Mean   : -10.78
##  3rd Qu.: 323.0   3rd Qu.: 545.0   3rd Qu.:  67.61   3rd Qu.:  17.50
##  Max.   : 583.0   Max.   : 694.0   Max.   : 153.55   Max.   : 149.40
##   yaw_dumbbell      total_accel_dumbbell gyros_dumbbell_x    gyros_dumbbell_y
##  Min.   :-150.871   Min.   : 0.00        Min.   :-204.0000   Min.   :-2.10000
##  1st Qu.: -77.644   1st Qu.: 4.00        1st Qu.:  -0.0300   1st Qu.:-0.14000
##  Median :  -3.324   Median :10.00        Median :   0.1300   Median : 0.03000
##  Mean   :   1.674   Mean   :13.72        Mean   :   0.1611   Mean   : 0.04606
##  3rd Qu.:  79.643   3rd Qu.:19.00        3rd Qu.:   0.3500   3rd Qu.: 0.21000
##  Max.   : 154.952   Max.   :58.00        Max.   :   2.2200   Max.   :52.00000
```

```
##  gyros_dumbbell_z  accel_dumbbell_x  accel_dumbbell_y  accel_dumbbell_z
##  Min.   : -2.380   Min.   :-419.00   Min.   :-189.00   Min.   :-334.00
##  1st Qu.: -0.310   1st Qu.: -50.00   1st Qu.:  -8.00   1st Qu.:-142.00
##  Median : -0.130   Median :  -8.00   Median :  41.50   Median :  -1.00
##  Mean   : -0.129   Mean   : -28.62   Mean   :  52.63   Mean   : -38.32
##  3rd Qu.:  0.030   3rd Qu.:  11.00   3rd Qu.: 111.00   3rd Qu.:  38.00
##  Max.   :317.000   Max.   : 235.00   Max.   : 315.00   Max.   : 318.00
##  magnet_dumbbell_x magnet_dumbbell_y magnet_dumbbell_z  roll_forearm
##  Min.   :-643.0    Min.   :-3600     Min.   :-262.00   Min.   :-180.0000
##  1st Qu.:-535.0    1st Qu.:  231     1st Qu.: -45.00   1st Qu.:  -0.7375
##  Median :-479.0    Median :  311     Median :  13.00   Median :  21.7000
##  Mean   :-328.5    Mean   :  221     Mean   :  46.05   Mean   :  33.8265
##  3rd Qu.:-304.0    3rd Qu.:  390     3rd Qu.:  95.00   3rd Qu.: 140.0000
##  Max.   : 592.0    Max.   :  633     Max.   : 452.00   Max.   : 180.0000
##  pitch_forearm     yaw_forearm      total_accel_forearm gyros_forearm_x
##  Min.   :-72.50   Min.   :-180.00   Min.   :  0.00      Min.   :-22.000
##  1st Qu.:  0.00   1st Qu.: -68.60   1st Qu.: 29.00      1st Qu.: -0.220
##  Median :  9.24   Median :   0.00   Median : 36.00      Median :  0.050
##  Mean   : 10.71   Mean   :  19.21   Mean   : 34.72      Mean   :  0.158
##  3rd Qu.: 28.40   3rd Qu.: 110.00   3rd Qu.: 41.00      3rd Qu.:  0.560
##  Max.   : 89.80   Max.   : 180.00   Max.   :108.00      Max.   :  3.970
##  gyros_forearm_y     gyros_forearm_z    accel_forearm_x   accel_forearm_y
##  Min.   : -7.02000   Min.   : -8.0900   Min.   :-498.00   Min.   :-632.0
##  1st Qu.: -1.46000   1st Qu.: -0.1800   1st Qu.:-178.00   1st Qu.:  57.0
##  Median :  0.03000   Median :  0.0800   Median : -57.00   Median : 201.0
##  Mean   :  0.07517   Mean   :  0.1512   Mean   : -61.65   Mean   : 163.7
##  3rd Qu.:  1.62000   3rd Qu.:  0.4900   3rd Qu.:  76.00   3rd Qu.: 312.0
##  Max.   :311.00000   Max.   :231.0000   Max.   : 477.00   Max.   : 923.0
##  accel_forearm_z   magnet_forearm_x  magnet_forearm_y magnet_forearm_z
##  Min.   :-446.00   Min.   :-1280.0   Min.   :-896.0   Min.   :-973.0
##  1st Qu.:-182.00   1st Qu.: -616.0   1st Qu.:   2.0   1st Qu.: 191.0
##  Median : -39.00   Median : -378.0   Median : 591.0   Median : 511.0
##  Mean   : -55.29   Mean   : -312.6   Mean   : 380.1   Mean   : 393.6
##  3rd Qu.:  26.00   3rd Qu.:  -73.0   3rd Qu.: 737.0   3rd Qu.: 653.0
##  Max.   : 291.00   Max.   :  672.0   Max.   :1480.0   Max.   :1090.0
##     classe
##  Length:19622
##  Class :character
##  Mode  :character
##
##
##
```

check if there is any remaining NA value

```
sum(!is.na(data3))
```

```
## [1] 1039966
```

read the tesint data and keep the same columns as cleaned training data

```
testing <- read.csv("pml-testing.csv")
test1 <- subset(testing,select = names(data3)[1:52])
str(test1)
```

```
## 'data.frame':    20 obs. of  52 variables:
##  $ roll_belt           : num  123 1.02 0.87 125 1.35 -5.92 1.2 0.43 0.93 114 ...
##  $ pitch_belt          : num  27 4.87 1.82 -41.6 3.33 1.59 4.44 4.15 6.72 22.4 ...
##  $ yaw_belt            : num  -4.75 -88.9 -88.5 162 -88.6 -87.7 -87.3 -88.5 -93.7 -13.1 ...
##  $ total_accel_belt    : int  20 4 5 17 3 4 4 4 4 18 ...
##  $ gyros_belt_x        : num  -0.5 -0.06 0.05 0.11 0.03 0.1 -0.06 -0.18 0.1 0.14 ...
##  $ gyros_belt_y        : num  -0.02 -0.02 0.02 0.11 0.02 0.05 0 -0.02 0 0.11 ...
##  $ gyros_belt_z        : num  -0.46 -0.07 0.03 -0.16 0 -0.13 0 -0.03 -0.02 -0.16 ...
##  $ accel_belt_x        : int  -38 -13 1 46 -8 -11 -14 -10 -15 -25 ...
##  $ accel_belt_y        : int  69 11 -1 45 4 -16 2 -2 1 63 ...
##  $ accel_belt_z        : int  -179 39 49 -156 27 38 35 42 32 -158 ...
##  $ magnet_belt_x       : int  -13 43 29 169 33 31 50 39 -6 10 ...
##  $ magnet_belt_y       : int  581 636 631 608 566 638 622 635 600 601 ...
##  $ magnet_belt_z       : int  -382 -309 -312 -304 -418 -291 -315 -305 -302 -330 ...
##  $ roll_arm            : num  40.7 0 0 -109 76.1 0 0 0 -137 -82.4 ...
##  $ pitch_arm           : num  -27.8 0 0 55 2.76 0 0 0 11.2 -63.8 ...
##  $ yaw_arm             : num  178 0 0 -142 102 0 0 0 -167 -75.3 ...
##  $ total_accel_arm     : int  10 38 44 25 29 14 15 22 34 32 ...
##  $ gyros_arm_x         : num  -1.65 -1.17 2.1 0.22 -1.96 0.02 2.36 -3.71 0.03 0.26 ...
##  $ gyros_arm_y         : num  0.48 0.85 -1.36 -0.51 0.79 0.05 -1.01 1.85 -0.02 -0.5 ...
##  $ gyros_arm_z         : num  -0.18 -0.43 1.13 0.92 -0.54 -0.07 0.89 -0.69 -0.02 0.79 ...
##  $ accel_arm_x         : int  16 -290 -341 -238 -197 -26 99 -98 -287 -301 ...
##  $ accel_arm_y         : int  38 215 245 -57 200 130 79 175 111 -42 ...
##  $ accel_arm_z         : int  93 -90 -87 6 -30 -19 -67 -78 -122 -80 ...
##  $ magnet_arm_x        : int  -326 -325 -264 -173 -170 396 702 535 -367 -420 ...
##  $ magnet_arm_y        : int  385 447 474 257 275 176 15 215 335 294 ...
##  $ magnet_arm_z        : int  481 434 413 633 617 516 217 385 520 493 ...
##  $ roll_dumbbell       : num  -17.7 54.5 57.1 43.1 -101.4 ...
##  $ pitch_dumbbell      : num  25 -53.7 -51.4 -30 -53.4 ...
##  $ yaw_dumbbell        : num  126.2 -75.5 -75.2 -103.3 -14.2 ...
##  $ total_accel_dumbbell: int  9 31 29 18 4 29 29 29 3 2 ...
##  $ gyros_dumbbell_x    : num  0.64 0.34 0.39 0.1 0.29 -0.59 0.34 0.37 0.03 0.42 ...
##  $ gyros_dumbbell_y    : num  0.06 0.05 0.14 -0.02 -0.47 0.8 0.16 0.14 -0.21 0.51 ...
##  $ gyros_dumbbell_z    : num  -0.61 -0.71 -0.34 0.05 -0.46 1.1 -0.23 -0.39 -0.21 -0.03 ...
##  $ accel_dumbbell_x    : int  21 -153 -141 -51 -18 -138 -145 -140 0 -7 ...
##  $ accel_dumbbell_y    : int  -15 155 155 72 -30 166 150 159 25 -20 ...
##  $ accel_dumbbell_z    : int  81 -205 -196 -148 -5 -186 -190 -191 9 7 ...
##  $ magnet_dumbbell_x   : int  523 -502 -506 -576 -424 -543 -484 -515 -519 -531 ...
##  $ magnet_dumbbell_y   : int  -528 388 349 238 252 262 354 350 348 321 ...
##  $ magnet_dumbbell_z   : int  -56 -36 41 53 312 96 97 53 -32 -164 ...
##  $ roll_forearm        : num  141 109 131 0 -176 150 155 -161 15.5 13.2 ...
##  $ pitch_forearm       : num  49.3 -17.6 -32.6 0 -2.16 1.46 34.5 43.6 -63.5 19.4 ...
##  $ yaw_forearm         : num  156 106 93 0 -47.9 89.7 152 -89.5 -139 -105 ...
##  $ total_accel_forearm : int  33 39 34 43 24 43 32 47 36 24 ...
##  $ gyros_forearm_x     : num  0.74 1.12 0.18 1.38 -0.75 -0.88 -0.53 0.63 0.03 0.02 ...
##  $ gyros_forearm_y     : num  -3.34 -2.78 -0.79 0.69 3.1 4.26 1.8 -0.74 0.02 0.13 ...
##  $ gyros_forearm_z     : num  -0.59 -0.18 0.28 1.8 0.8 1.35 0.75 0.49 -0.02 -0.07 ...
```

```
##  $ accel_forearm_x     : int  -110 212 154 -92 131 230 -192 -151 195 -212 ...
##  $ accel_forearm_y     : int  267 297 271 406 -93 322 170 -331 204 98 ...
##  $ accel_forearm_z     : int  -149 -118 -129 -39 172 -144 -175 -282 -217 -7 ...
##  $ magnet_forearm_x    : int  -714 -237 -51 -233 375 -300 -678 -109 0 -403 ...
##  $ magnet_forearm_y    : int  419 791 698 783 -787 800 284 -619 652 723 ...
##  $ magnet_forearm_z    : int  617 873 783 521 91 884 585 -32 469 512 ...
```

```
sum(is.na(test1))
```

```
## [1] 0
```

data preparation: training split into training and testing

```
library(caret)
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
inTrain <- createDataPartition(y=data3$classe,p=0.7,list=FALSE)
training <- data3[inTrain,]
testing <- data3[-inTrain,]
```

decisition tree

```
treefit <- train(classe~.,data=training,method="rpart")
predtree <- predict(treefit,testing)
table(predtree,testing$classe)
```

```
##
## predtree    A    B    C    D    E
##        A 1020  157   30   48   15
##        B    4  198   20    5    5
##        C  305  194  679  303  209
##        D  316  590  297  608  348
##        E   29    0    0    0  505
```

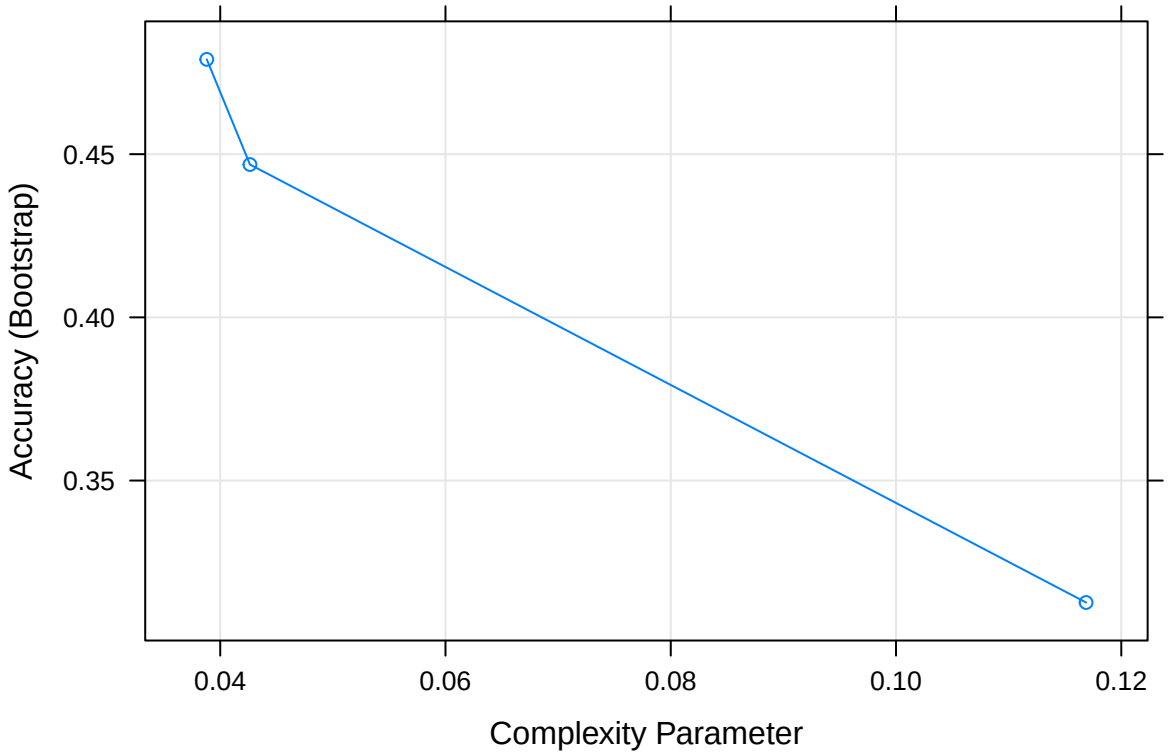
```
confusionMatrix(predtree,factor(testing$classe))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1020  157   30   48   15
##          B    4  198   20    5    5
##          C  305  194  679  303  209
##          D  316  590  297  608  348
```

```
##          E   29    0    0    0  505
##
## Overall Statistics
##
##                Accuracy : 0.5115
##                  95% CI : (0.4986, 0.5243)
##     No Information Rate : 0.2845
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.3925
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.6093  0.17384   0.6618   0.6307  0.46673
## Specificity            0.9406  0.99284   0.7919   0.6848  0.99396
## Pos Pred Value         0.8031  0.85345   0.4018   0.2816  0.94569
## Neg Pred Value         0.8583  0.83354   0.9173   0.9045  0.89217
## Prevalence             0.2845  0.19354   0.1743   0.1638  0.18386
## Detection Rate         0.1733  0.03364   0.1154   0.1033  0.08581
## Detection Prevalence   0.2158  0.03942   0.2872   0.3669  0.09074
## Balanced Accuracy      0.7750  0.58334   0.7269   0.6578  0.73035
```

Plot decision tree

```
plot(treefit)
```



Random forest

```
#modFit <- train(classe~.,data=training,method="rf",prox=TRUE)
#modFit
rfFit <- readRDS("randomforest.rds")
predrf <- predict(rfFit,testing)
table(predrf,testing$classe)
```

```
##
## predrf    A    B    C    D    E
##      A 1673    2    0    0    0
##      B    1 1137    1    0    0
##      C    0    0 1024    2    2
##      D    0    0    1  962    3
##      E    0    0    0    0 1077
```

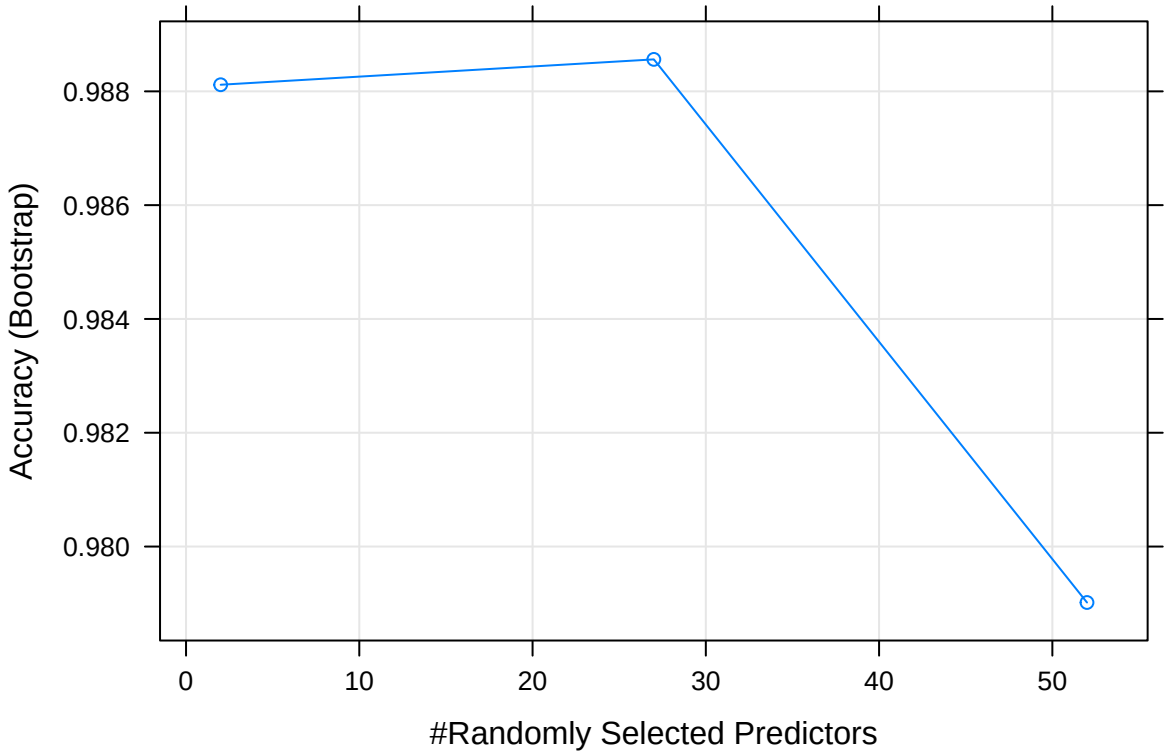
```
confusionMatrix(predrf,factor(testing$classe))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1673    2    0    0    0
```

```
##          B    1 1137    1    0    0
##          C    0    0 1024    2    2
##          D    0    0    1  962    3
##          E    0    0    0    0 1077
##
## Overall Statistics
##
##                Accuracy : 0.998
##                  95% CI : (0.9964, 0.9989)
##     No Information Rate : 0.2845
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9974
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.9994   0.9982   0.9981   0.9979   0.9954
## Specificity            0.9995   0.9996   0.9992   0.9992   1.0000
## Pos Pred Value         0.9988   0.9982   0.9961   0.9959   1.0000
## Neg Pred Value         0.9998   0.9996   0.9996   0.9996   0.9990
## Prevalence             0.2845   0.1935   0.1743   0.1638   0.1839
## Detection Rate         0.2843   0.1932   0.1740   0.1635   0.1830
## Detection Prevalence   0.2846   0.1935   0.1747   0.1641   0.1830
## Balanced Accuracy      0.9995   0.9989   0.9986   0.9986   0.9977
```

Plot RF model

```
plot(rfFit)
```



gradient Boosted trees

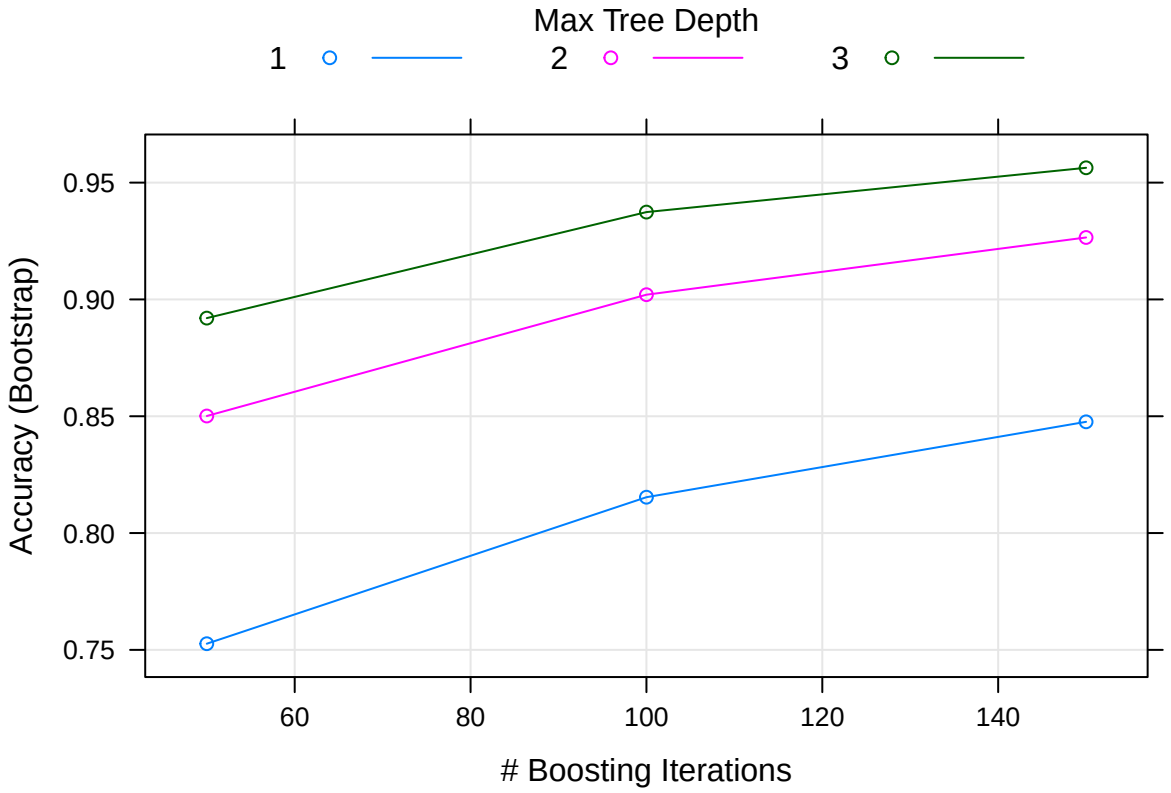
```
#mod_gbm <- train(classe~.,data=training,method="gbm",verbose=F)
mod_gbt <- readRDS("mod_gbm")
predgbm <- predict(mod_gbt,testing)
confusionMatrix(predgbm,factor(testing$classe))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1656   30    0    2    0
##          B   11 1094   23    3   13
##          C    5   15  992   31    8
##          D    1    0    7  922   11
##          E    1    0    4    6 1050
##
## Overall Statistics
##
##                Accuracy : 0.9709
##                  95% CI : (0.9663, 0.9751)
##     No Information Rate : 0.2845
##     P-Value [Acc > NIR] : < 2.2e-16
```

```
##
##                   Kappa : 0.9632
##
##  Mcnemar's Test P-Value : 1.914e-07
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.9892   0.9605   0.9669   0.9564   0.9704
## Specificity            0.9924   0.9895   0.9879   0.9961   0.9977
## Pos Pred Value         0.9810   0.9563   0.9439   0.9798   0.9896
## Neg Pred Value         0.9957   0.9905   0.9930   0.9915   0.9934
## Prevalence             0.2845   0.1935   0.1743   0.1638   0.1839
## Detection Rate         0.2814   0.1859   0.1686   0.1567   0.1784
## Detection Prevalence   0.2868   0.1944   0.1786   0.1599   0.1803
## Balanced Accuracy      0.9908   0.9750   0.9774   0.9763   0.9841
```

Plot GBT

```
plot(mod_gbt)
```



Conclusion

The best fit model is randow forest:with 0.9976 accuracy outperforming other models

prediction on the test data set

```
predtest <- predict(rfFit,test1)
predtest
```

```
##  [1] B A B A A E D B A A B C B A E E A B B B
## Levels: A B C D E
```